

統計建模、機器學習與決策制定

莊皓鈞*

我於 2013 年自 Texas A&M University 的 Mays 商學院取得作業與供應鏈管理博士學位（副修：應用統計），返國於國立政治大學任教至今，由於求學期間接受了大量的計量經濟、統計分析、電腦模擬、管理科學方法論訓練，所做研究多為針對產業營運問題建構對應數據分析與量化模型，以期協助決策優化並產出學理貢獻。涉獵的產業跨足消費零售、線上平臺、電子製造，歷經鑽研這些議題十數載，以及投稿期刊過程中反覆修正的磨練，體悟到即便數據型態與問題方向不同，始終不脫**統計建模**（statistical modeling）、**機器學習**（machine learning）與**決策制定**（decision making）這三者的集合，而這三者也正是當代方興未艾的大數據分析、人工智慧的核心。然而在研究發表到實際應用的過程中，卻深刻感受到三者時有未能理想整合，造成了理論與實務脫鉤，因此想藉此一短文，與同儕、先進分享交流累積的心得與反思。

在運用量化數據進行實證分析與研究問題解決過程中，傳統統計建模的訓練讓學者習於配適模型——目標變數 $=f(\text{參數} * \text{預測變數})$ ，用於理論／假說測試。在應用計量經濟／統計方法時，管理實證學者常習於先假定 $f(\cdot)$ 為一線性可連加函式，延用估計方法力求能夠產生不偏、一致與有效率參數估計值，接著仰賴估計值信賴區間和 p 值來達成研究結論。這樣的研究典範，在我生涯初期（2009-2014）習以為常，數據分析設法樣本估計值能近似未知母體真實參數，以最大化實證效度，並透過參數估計值來瞭解預測變數 x 與目標變數 y 的關聯／因果關係，格外關注統計上的顯著性、樣本的模型配適度，以及理論計量經濟／統計推論是否有更堅實的估計方式。

上述以統計理論為基礎產出最佳不偏線性估計模型 $f(\cdot)$ ，以測試理論的研究思維，在 2015 年後受到機器學習很大的衝擊。隨著預測分析（predictive analytics）與諸多機器學習演算法隨著 Python、R 等開源運算分析軟體的擴散採用，我接觸到截然不同有別早期統計／計量訓練的思維，特別有三點：（一）除

* 國立政治大學資訊管理學系副教授

樣本配適程度 (in-sample fitness) 外，樣本外的預測能力 (out-of-sample prediction accuracy) 與透過驗證樣本調整模型參數／架構成了數據分析重心。(二) 模型函式 $f()$ 的線性假設與參數估計的統計推論非首要的關切重點，估計方法為含高度非線性轉換能力的演算法，如類神經網路、隨機森林、梯度提升機、支持向量機等。(三) 訓練的模型不再必然能讓研究者直觀瞭解預測變數 x 與目標變數 y 的關聯顯著和效果大小，在不清楚模型 $f()$ 的實證意義下，追求的是精準預測目標 y 。

而這預測分析浪潮，隨著深度學習 (deep learning) 興起，越演越烈。我在此浪潮下以數理統計、最佳化理論根基學習機器學習並將其應用於研究並未過於困難。然而就社會科學與管理研究而言，提供理論解釋 (explanation) 以概化母體，相較機器學習力求準確預測 (prediction) 捕捉異質個體，研究思維大相逕庭。2015 年後約二至三年，許多實務問題激發的研究問題皆在追求建構模型 $f()$ 以最小化目標變數 y 的預測誤差目標，我初期習以為常的統計建模與顯著測試思維也被研究的合作企業／夥伴擱置，即便發表了數篇預測分析文章，因自身來自統計建模實證研究典範，對傳統所仰賴的線性模型 $f()$ 成了高度非線性的黑箱模型，且參數估計值是否有良好統計性質 (不偏、一致、有效率) 未能確知，我時常感到困惑與苦惱。

而這樣的解釋性統計建模 vs 預測性機器學習二元差異甚至是衝突，有幸在近三年 (2018-) 的研究歷程中，隨著實務需求與理論研究的更新，得到了緩解與新知。首先是實務夥伴在得到複雜但相較於傳統線性模型能改善預測準度的模型後，十分關注究竟哪些輸入預測變數 x 重要且顯著？預測因子 x 之於目標變數 y 的影響／關係是如何？產出對目標變數 y 的點預測估計偏誤為何？是否有統計上可靠的區間估計？這些問題，經歷了一陣精準預測研究熱潮後，回到了傳統統計建模的範疇，並與興起的可詮釋 (interpretable) 機器學習需求息息相關。在統計與計量經濟領域，晚近的理論研究也發展出以隨機抽樣為基礎的估計值標準誤、信賴區間、顯著性檢定等實證量化研究基礎工具，將統計建模與機器學習兩者共通的脈絡逐步建構起來，我也在其中找到了整合兩者研究的機會，得到了許多想法的啟迪與經驗，逐漸掌握到作為管理與社會科學研究者，如何在技術本位的機器學習浪潮，創造出研究的價值，以下野人獻曝分享三個與產業合作經驗。

第一個案例是與某間全球位列前端的半導體電子零組件通路商合作，進行供應鏈中的工業客戶需求預測，由於需求資料本身為時間序列數據，很自然地一開始就嘗試以時間序列預測模型 (ARIMA、Exponential Smoothing) 及更複雜

的類神經網路模型處理此問題。意外地，深度類神經網路和時間序列模型準確度卻由於過度擬合 (over-fitting) 表現遜於企業自身的簡單移動平均模型，對當時初入機器學習領域的我，甚感震撼。爾後，深入思考此預測問題在管理營運面上的本質，將資料由高度不穩定的週需求，聚合成與預測目標 (12 週總需求) 對應的季需求量顆粒度，並運用作為作業管理學者的需求預測理論知識，從需求數據建構出具營運意義的預測變數。最後運用追蹤資料的估計概念，將數據進行跨品項學習 (cross-item learning) 配適一個整體預測模型。有別於前述個別品項的時間序列模型表現不佳，此模型大幅改善預測準確度，研究成果發表於領域頂尖期刊 *Journal of Operations Management*。此案例始於不成功的深度學習，結於運用理論知識進行資料處理與變數建構，達成成功的整體機器學習，讓我體會到過往實證研究的思維訓練，對於發揮機器學習價值甚有助益，甚者，也領略了直接套用演算法於數據，難以改善研究目標。

第二個案例是與線上平臺業者合作，研究如何運用平臺用戶的活動及消費數據，建立以機器學習為基礎的回訪預測模型，目標為針對每一個用戶，產出其在未來某一特定時間區段 (如週、月、季) 是否會回訪。在早年資料探勘 (data mining) 興起時，此類問題即有為數眾多的相關研究，近年來又因機器學習而蓬勃發展。研究進行時，平臺有著出色的工程團隊，建構數種尖端的監督式機器學習預測模型。然而瓶頸在於許多演算法都是針對橫斷面 (cross-sectional) 資料設計，其觀察值間獨立性假設 (iid) 並不合於平臺的用戶追蹤資料數據；另用高維度的客戶特徵變數為預測因子時，無法清楚瞭解因子 x 與預測目標變數 y 之間的效果和關係。我由計量經濟理論文獻得知 Lasso 迴歸模型在追蹤資料仍具有良好統計效度，且 Lasso 專長於最佳化將影響力低的預測變數 x 自模型中排除，正好十分符合研究的理論與實務需求。從平臺數據的實證分析，也發現清楚易懂的 Lasso 迴歸模型，預測準確度甚至勝過複雜的循環式神經網路、長短期記憶模型等，研究成果發表於領域一流期刊 *European Journal of Operational Research*。研究過程中，過往統計／計量經濟迴歸建模的訓練基礎，引導我從統計學習觀點來切入典型的機器學習問題，形成學術與實務貢獻。後續一些研究受此案例激發，更深入的整合統計方法與機器學習，異中求同，嘗試創新。

第三個案例是與工研院巨量資料中心合作，研究如何在零售業估計需求的行銷效應。合作方有著許多受過資訊科學良好訓練的工程團隊，倘若只是要建構機器學習模型降低銷售量／金額的預測錯誤，我作為商管學者能夠提供的價值十分有限。然而合作方感到困擾欲分析的議題，是零售業者針對不同的品類

在不同時期，實施降價促銷、廣告宣傳、特殊展示等行銷方案，卻難以估計這些方案對品類自身銷售的效果，以及行銷方案對其他品類銷售可能的帶動互補或干擾替代效果。要解決這個數據分析問題，模型除了需產出清楚的行銷效果估計值，還要能考慮到品類銷售是個高維度時間序列系統，也就是各個品類都可視為個體，其銷售狀況隨時間波動外，彼此也存在互相帶動或干擾替代的隱藏連動關係。在這個研究案中，過往所受的計量經濟訓練，讓我自然地找到了經典的向量自我迴歸（VAR）作為起點，然而大量零售銷售品類及行銷方案變數形成的高維度 VAR 系統，傳統計量理論並無涵蓋。此時第二個案例所累積的 Lasso 技術使用經驗，幫助我設計出 VAR-Lasso 的分析方法，成功應用在實際大量零售數據回答了複雜的研究議題。這個研究除了讓我自單一目標變數模型，跨入更複雜的高維度變數系統模型，更再一次強化了我整合統計／計量經濟理論與機器學習技術能產出學術與實務價值的信念。

上述的研究歷程大抵在建構出有效的模型目標變數 $=f(\text{參數} * \text{預測變數})$ 後告一段落，也成功發表於優質國際學術期刊，然而作為一位「入世」嘗試以研究解決產業問題的管理學者，卻發現統計建模與機器學習在實務上應用的鴻溝，也就是**決策制定**的層面。上列兩個案例，客戶需求預測的目的是幫助企業能做出更好的採購庫存規劃；用戶回訪預測的目的是讓平臺能更好的運用行銷資源設定活動標的。一般而言，再好的模型都會有相當程度的預測錯誤，所產出的預測值多是企業決策流程或個體決策模型中的一個參數。從實證研究產出決策／政策處方（prescription），重點並非模型 $f(\)$ 或是參數估計值，在於優化輸入因子中可控制的部分，進而最大或最小化決策情境目標（成本、效益）。當數據分析止步於描述、解釋或預測，而無產出決策制定的處方時，也容易造成實務上組織投入資源蒐集整理資料，針對標的進行統計建模、機器學習產出模型後，卻感受不到模型對於營運與決策的影響，連帶使產業對於數據分析價值存疑。

上述數據分析、實證研究的價值困境，對應到近期的一種研究思維，從先預測後決策（predict then prescribe），改為預測即決策（predict and prescribe）。概念上也就是將決策當成預測的標的，透過統計建模與機器學習，從數據學習輸入預測因子和最佳決策方案的關係，我曾於過往與一家連鎖高端女性服飾通路業者合作中，做過此類型研究。簡單來說，服飾通路商如同許多零售、服務業者，日常的營運決策必有店內的人力配置，過往研究思維多為先產出每日的銷售或來客數預測，再依此預測值設定每時段人力數。此做法的缺陷在於銷售／來客預測不可避免的會有誤差，將連帶造成後續的人力決策處方誤差（prescription error）。而我在此案例研究手法是以數據估計銷售量之於人力、來客數的反應函

數後，以計算出的最佳人力數目作為預測標的，免去產出第一階段預測，達到預測即決策。後續以實證資料驗證此理論方法的利潤表現，十分靠近有完美來客數資訊下的最佳利潤，將此研究發表於領域頂尖期刊 *Production and Operations Management*，並優化營運實務。

這些逐年匯集的研究案例經驗，反映管理實證研究在大數據分析、人工智慧下，固然研究發表可能來自統計建模、機器學習、決策制定其中單一領域，然而以數據建構模式後，要能發揮實證分析的價值，三個領域實是環環相扣，需有整體性思維看待這些在實務和學術上時常未能整合的面向。最後，人工智慧發展實已有數十載，機器學習領域的演進和更多數據的可得性雖使人工智慧得以開創新局，其中仍有許多設計、架構面的問題是需要領域理論知識來解決。作為非機器學習理論的應用學者，一方面投入時間學習新的方法，另一方面終究需回到研究領域根本的核心問題，思考這些技術、演算法的突破如何能創造價值。這些方法論創新就像是把鐵鎚，然而很多深刻的問題，並不如釘子一槌即可中的，演算法不足以創造見解。我也仍持續從未符預期的錯誤與不理想的研究結果中，苦思、勤學，以求最終得到嶄新知識，對文獻與實務期盼有些許貢獻，共勉之。